Lexis
Journal in English Lexicology

5 | 2010
Lexicology & Stylistics

A Corpus-Based Case Study in Prepositional AT/ TO- Infinitive Alternation using the Lemma "AIM"

Kate James

Electronic version
URL: http://journals.openedition.org/lexis/505
DOI: 10.4000/lexis.505
ISSN: 1951-6215

Publisher
Université Jean Moulin - Lyon 3

Electronic reference
Kate James, « A Corpus-Based Case Study in Prepositional AT/TO- Infinitive Alternation using the Lemma "AIM" », *Lexis* [Online], 5 | 2010, Online since 15 October 2010, connection on 02 May 2019. URL : http://journals.openedition.org/lexis/505 ; DOI : 10.4000/lexis.505

A Corpus-Based Case Study in Prepositional AT/TO- Infinitive Alternation using the Lemma "AIM"

Kate James[1]

Abstract

The lemma 'AIM' in its verbal form presents a case of prepositional (*at*)/to-infinitive alternation. This study demonstrates that whilst users of monolingual and bilingual dictionaries would conclude that the pair is equivalent in use, a corpus study of the two forms enables certain clear distinctions to be made. The most significant of these may be found to be the predominant pattern of use of *AIM+at* in more formal styles with *AIM+to* being prefered in less formal discourse. Other uses are examined in context concerning clausal polarity, tense, and collocational elements of semantic significance. The relevance of such a corpus-based study is examined in relation to possible use in the field of stylistics. This study concludes by demonstrating the importance of corpus study to supplement information given in dictionaries, and suggesting the highlighting of distinctions in underlying principles through the use of such data, as well as noting possible relevance to the stylistician.

Keywords: *aim to* – *aim at* – corpus – lexicology – stylistics – discourse

***

Résumé

L'item lexical verbal 'AIM' présente une alternance prépositionnelle dans sa forme infinitive (*at/to*). Les utilisateurs de dictionnaires mono- ou bilingues pourraient conclure que les deux formes sont utilisées de façon identique. Cependant, cette étude essaie de démontrer que des différences bien identifiables existent. Une des plus significatives concerne l'utilisation de *AIM + at* dans un contexte formel, alors que l'utilisation de *AIM + to* semble prédominer dans une situation plus informelle. D'autres utilisations sont examinées dans leur contexte concernant la polarité clausale, le temps, et des éléments contextuels qui comportent une importance sémantique. La pertinence de cette étude appliquée au domaine de la stylistique est également examinée. En conclusion, cette étude démontrera l'importance des études de corpus afin de fournir des informations supplémentaires dans les définitions de dictionnaires. Elle notera également l'importance qu'il y a à différencier les principes lexicaux fondamentaux par l'utilisation de ces données. La pertinence de ces données pour le stylisticien sera discutée.

Mots-clés : *aim to* – *aim at* – corpus – lexicologie – stylistique – registre de langue

[1] Université d'Auvergne, IUT, 100 rue de l'Egalité, 15000 Aurillac : kate.james@iut.u-clermont1.fr

Introduction

Traditionally, the corpus-driven lexicographer is concerned with determining the meaning and pattern of words through large-scale data analysis, whereas the stylistician can be seen to concentrate on the variability of linguistic forms in a text and the resulting connotations and artistic effects thus created. Both, however, can be described as using empirical approaches, whether based on language usage determined through corpora analysis, or linguistic evidence in the text, and as pointed out by Mahlberg [2006]:

> [B]oth are interested in HOW we say what we say.

Thus, some overlap may be seen in corpus-based lexical and stylistic investigations. As Stubbs [2005] points out:

> Individual texts can be explained only against a background of what is normal and expected in general language use, and this is precisely the comparative information that quantative corpus data can provide.

As well as the more usual questions of a semantic nature, grammatical features may also interest the lexicographer, keen to distinguish patterns of usage. This may also be of concern to the stylistician, desirous to identify inherent stylistic qualities of a text. In this paper, a study of preposition choice following the lemma 'AIM' will be examined from a corpus-based viewpoint for lexicological purposes. Attempts will then be made to determine relevance to stylistics.

The lemma 'AIM' in its verbal form presents a case of prepositional (*at*)/*to*-infinitive alternation. Faced with such a choice, the user or learner of English may confront certain difficulties as to the use of the "correct" form, and seek to discover patterns of differing use based on a variety of sources. According to Sinclair [1991: 37], the three main sources of lexicographic evidence listed in order of popularity are:

1. dictionaries
2. [native speaker] users' ideas about their language
3. observation of language in use

These categories, based on a lexicographer's study of the language, may mirror the procedure adopted by a learner seeking to justify the choice of one word form rather than the other although the accessibility or value of each source will vary. As Stubbs [2000] points out, defending the value of observing language in use by means of corpora in such a search:

> Native speakers have only very poor intuitions about many pragmatic aspects of language. If they had better intuitions then these pragmatic aspects would be recorded in dictionaries, but they often aren't. Corpus studies show up patterns which are not visible in single examples...Corpus linguistics is based on methods of observation which make repetitions visible, and the fact of repetition makes quantitative methods essential.

Echoing these concerns, Aarts [1991: 46] states that intuitive grammar, when combined with a corpus study, becomes an observation-based grammar. Structures such as 'AIM' *at/to* may necessitate such a study, described by Itkonen [1980: 344] as being suited to

> unclear cases, i.e. less than well-established rules, as well as actual linguistic behaviour in general, [which] must be investigated by means of observation.

If the observation of language allows certain rules of use to be established, this may also be of significance to the stylistician.

1. Corpus use

Although modern dictionaries are corpus-based, Stubbs [2000] points out that

> patterns are not explicitly recorded [...] The examples in the dictionary confirm [corpus] data, but the underlying principles are not given.

This may imply that only a corpus study of a given word will enable correct patterns to be found by comparing data using quantitative techniques. This idea is supported by Aarts [1988] who claims that

> only linguists who use corpus data themselves will know that a corpus yields a much greater variety of constructions than one can either find in the literature or think up oneself.

Using corpora, the great quantity and variety of constructions may be accessed and visualised as a whole and, as Sinclair [1991: 100] states,

> language looks rather different when you look at a lot of it at once.

The corpus used in this study is the smaller collection of corpora from the Bank of English (BoE), the CobuildDirect (CobuildBoE) corpus comprising 56 million words. The individual corpora which make up this corpus may be found in Appendix 1. The corpus was accessed using Telnet and data collected using COBUILD corpus access software (CAS). After loading saved files for home use, certain data were further treated with VisualBasic programming using Microsoft Excel software for ease of access.

2. Corpus analysis

2.1. Corpus query

2.1.1. Query form

The data analysed were obtained using the queries **aim@+0,2at+VERB** and **aim@+0,2to+VERB**. The results obtained were then further selected by retaining only the verbal forms of *AIM*, thus rejecting the nominal usages *aim* and *aims*. As corpus tagging is automatic, accuracy may not be total [Potter 1999: 35] and errors are thus likely be found; indeed, Kennedy estimates the margin of error as being 3% to 4% [Kennedy 1998: 212].

In order to check the accuracy of the data obtained for this study, a manual verification was carried out on the data bearing verbal tagging as well as the non-retained data with nominal tagging. No obvious errors were noted concerning data obtained for verbal group tagging whereas in data selected for nominal tagging, 35 cases of the verbal form of *AIM* were

found. The incorporation of these omissions to collected data allowed a slight error margin of 1.7% to be corrected. The errors are consistent with De Roses's identification of the five most frequent tagging errors concerning, in this case, singular nouns assigned instead of a verb [De Rose 1991: 9-14]. Figure 1 shows the number of examples found in the corpus as well as in each sub-corpus, expressing in each case the estimated number of occurrences found per million words of text. 843 cases of *AIM at* and 1195 instances of *AIM to* can thus be found.

Corpus	**aim@+0,2at+VERB**	
	Number of occurrences	**Average/ million words**
bbc	207	7
oznews	167	31.3
npr	115	36.8
times	100	17.3
today	59	11.2
sunnow	47	8.1
ukbooks	39	7.3
ukmags	34	6.9
ukephem	33	10.6
usbooks	27	4.8
ukspok	9	1.0
usephem	6	4.9
TOTAL	**843**	

Corpus	**aim@+0,2to+VERB**	
	Number of occurrences	**Average/ million words**
sunnow	218	37.
ukephem	177	56.7
times	171	29.7
today	167	31.8
ukmags	141	28.8
oznews	113	21.2
ukbooks	74	13.8
bbc	52	19.9
usbooks	37	6.6
ukspok	28	3.0
npr	11	3.5
usephem	6	4.9
TOTAL	**1195**	

Figure 1: Frequency of occurrence and average frequency per million words of *AIM at* and *AIM to*

It may be noted that the average number of occurrences changes, sometimes considerably, depending on the sub-corpus examined and on the form requested. Possible reasons for this difference will be proposed in 3.2.

2.1.2. Distance span

The corpus query was defined in order to allow the inclusion of sentences with two words occurring between the lemma *AIM* and *to* or *at.* This distance was determined by changing the distance values of the original query, starting with **0,0** and continuing up to **0,5**. The optimal distance span for *AIM at* was found to be **0,2,** the number of relevant cases steadying after this value. This was also found to be a reasonable distance span for *AIM to*, evoking a maximum of relevant concordances with a minimum of non-relevant ones, only two occurrences being excluded. As *as* and *to* complement the meaning of *AIM*, no relevant cases seem likely after a span of **0,5**.

Results show that both *AIM to* and *AIM at* are separated by intervening words in relatively few cases (seventeen times with *AIM at* and twelve times with *AIM to)*, that most cases involve quantifiers or adverbs and that very few are used more than once in the data collected. The only co-occurrence pattern which is significantly repeated is *AIM only/not only at* which occurs six times. However, the small proportion of occurrences with intermittent words does not on its own allow conclusions to be made as to possible differences of use concerning *AIM to* and *AIM at.*

2.2. Lexical profile

The study of *AIM at/to* separated from its surrounding context would give little results or, as Sinclair [1996] states

> the idea of a word carrying meaning of its own [can] be relegated to the margins of linguistic interest.

Stubbs [2001: 63] develops this idea stating that

> meaning is typically dispersed over several word-forms which habitually co-occur in text […] [and] these co-occurring word forms 'share' semantic features.

The lemma *AIM at/to,* as well as having an individual semantic definition, may also be seen as being strongly attached to its surrounding context and as such could be considered part of an extended unit of meaning as defined by Sinclair [1996].

The analysis of *AIM at/to* will be based on Sinclair's model of extended lexical units [1996] & [1998] as developed by Stubbs [2001: 87], who specifies possible constituents of the unit and the relation between them. Both aspects will be analysed using corpus data according to the aforementioned model, given as follows:

1) collocation
2) colligation
3) semantic preference
4) discourse prosody
5) strength of attraction
6) position and positional mobility
7) distribution in text types

As Sinclair [1998] mentions, the first four categories are increasingly abstract and according to Stubbs [2001: 88] are all "probabilistic and non directional", thus needing to be complemented by the last three relations which give more concrete statistical data and define the limits of application. Collocation and semantic preference seem here to be closely related and will be treated together. The strength of attraction, defined by Stubbs [2001: 88] as being "the probability of occurrence of a collocate, grammatical category, lexical set or discourse prosody" given the occurrence of a node will be discussed in each individual part of this study, statistical percentages being given according to actual occurrence.

2.2.1. Collocation and semantic preference

Sinclair [1991: 170] describes collocation as 'the occurrence of two or more words within a short space of each other in a text' and has further noted that the probabilities of lexical items are affected mainly within a span of around four [Sinclair *et al.* 1970] & [Sinclair 1985].

Looking at collocations occurring directly to the left of the node, it can be seen that the choice of lexical item tends to vary according to the subsequent use of *AIM to* or *AIM at.* Comparing the t score pictures in figure 2, it may be noted that the subsequent use of *AIM at* seems to result from the choice of more precise nominal subjects, such as *talks, campaign, measure* or *strategy,* thus reflecting the possibly more journalistic or formal tone of the

surrounding text. This point will be examined in more detail when considering the distribution in text types. The t scores of nouns occurring in the first twenty collocates of the node at N-1 are also given in figure 3. It may be noted that *AIM at* is not only preceded by a significantly higher number of nominal occurrences but also that the t-score is always greater than 2, a value that Barnbrook [1996: 98] considers to be indicative of the most interesting cases of collocation using this method of calculation. The immediate left collocations (N-1) of *AIM to* contain more grammatical words, such as *which* or quantifiers; both aspects shall be studied in 3.2.2. when considering colligation, which may here be seen as the grammatical company a word keeps, or avoids keeping, either within its own group or at a higher rank.

AIM at

a	a	is	NODE	at	ending	the
the	new	talks	NODE	only	reducing	a
of	which	was	NODE	particular	improving	out
announced	move	are	NODE	solely	getting	an
talks	spending	campaign	NODE	deliberate	helping	up
new	talks	were	NODE	specifical	bringing	its
series	visit	measures	NODE	mainly	resolving	difference
launched	peace	be	NODE	directly	forcing	people
an	bill	strategy	NODE	less	preventing	them
in	economic	plan	NODE	not	curbing	relations
tough	week	package	NODE	<f>	boosting	computer
operation	advertisin	policy	NODE	more	giving	iraq
conference	un	laws	NODE	in	increasing	young
meeting	latest	course	NODE	it	making	president
round	training	initiative	NODE	protecting	tensions	
by	plan	project	NODE	securing	tension	
week	special	programme	NODE	establishi	criticism	
monitor	yesterday	been	NODE	providing	violence	
welcome	intensive	today	NODE	creating	students	
scheme	proposals	initiative	NODE	developing	parents	
believed	policies	proposals	NODE	finding	down	
visit	would	programmes	NODE	keeping	public	
these	project	efforts	NODE	at	both	

AIM to

p>	we	we	NODE	to	make	the
said	<p>	which	NODE	continuous	help	a
season	the	he	NODE	successful	provide	up
</h>	this	is	NODE	generally	raise	them
new	<h>	it	NODE	eventually	give	an
<h>	now	will	NODE	therefore	get	back
boss	london	are	NODE	instead	keep	out
national	programme	should	NODE	above	build	pound
the	will	also	NODE	low	cut	students
now	which	they	NODE	further	show	its
unit	i	i	NODE	at	reduce	his
trade	and	company	NODE	also	put	it
lives	group	course	NODE	them	bring	all
says	jason	project	NODE	</h>	promote	together
year	how	now	NODE	not	achieve	each
war		m	NODE	one	develop	awareness
south	scheme	would	NODE	for	improve	improve

this	tour	degree	NODE	he	offer	people
1998	campaign	programme	NODE	in	become	scotland
preparator	bill	book	NODE	<p>	win	parents
hindu	labour	re	NODE	end	your	
maintenanc	start	always	NODE	prove	local	
raising	service	labour	NODE	attract	within	

Figure 2: t score pictures for the queries aim@+0,2at+VERB and aim@+0,2to+VERB

aim@+0,2at+VERB		
N-1	**t-score**	**Number of occurrences**
talks	7.13	51
campaign	3.97	16
measures	3.73	14
strategy	3.45	12
plan	3.12	10
package	2.99	9
policy	2.78	8
laws	2.63	7
course	2.51	7
initiative	2.44	6
project	2.42	6
programme	2.40	6

aim@+0,2to+VERB		
N-1	**t-score**	**Number of occurrences**
company	3.19	11
course	3.16	11
project	2.96	9
degree	2.41	6
programme	2.38	6
book	2.34	6

Figure 3: t scores and number of actual occurrences of nouns appearing in the top 20 N-1 collocates of *AIM at* and *AIM to*

A study of verbs falling directly to the right (N+1) of *AIM at* reveals that a total of 286 different verbal types are used with 116 being used more than once. With *AIM to*, these figures are 407 and 168 respectively. The collocations representing more than 1% of all usage at N+1 are given in figure 4 in order of frequency. This indicates that a wide semantic choice is available and the study of each individual case would be unlikely to show general patterns. However, when comparing the most popular collocations of both forms of the lemma, it can be seen that *AIM to* tends to favour the use of verbs possessing less semantic precision such as *MAKE, GET, PUT*. These are often used as phrasal verbs (*make up, get back, put up*) or as part of idiomatic use (*make the most of it, get a crack at, put it right*), implying text types of low perceived formality. The occurrence of these three verbs is lower following *AIM at.* Greater semantic precision may once again be detected in N+1 collocates of *AIM at,* reflecting observations made concerning nouns at N-1.

AIM at		
N+1 collocate	**Number of occurrences**	**% of usage as collocate**
ending	50	5.93
improving	29	3.44
reducing	29	3.44
helping	25	2.97
getting	24	2.85
bringing	19	2.25
forcing	17	2.02

AIM to		
N+1 collocate	**Number of occurrences**	**% of usage as collocate**
make	47	3.93
help	37	3.10
provide	37	3.10
get	26	2.18
raise	26	2.18
give	25	2.09
keep	20	1.67

resolving	17	2.02
preventing	15	1.78
curbing	14	1.66
boosting	13	1.54
giving	13	1.54
increasing	12	1.42
making	12	1.42
protecting	11	1.30
providing	11	1.30
establishing	10	1.19
finding	10	1.19
securing	10	1.19
creating	9	1.07
developing	9	1.07
keeping	9	1.07

build	18	1.51
improve	18	1.51
cut	17	1.42
put	17	1.42
bring	16	1.34
show	16	1.34
develop	15	1.26
become	14	1.17
end	14	1.17
promote	14	1.17
win	14	1.17
offer	13	1.09
prove	13	1.09
reduce	13	1.09
attract	12	1.00

Figure 4: N+1 collocation of aim@+0,2at+VERB and aim@+0,2to+VERB

Although a study of individual collocates does not seem pertinent, semantic preference may be seen as being significant, notably through various synonyms conveying the idea of increase, reduction and halting. Corresponding verbs of these denotational synonyms placed at N+1 are shown in figure 5. These account for 22.8% of collocates of *AIM at* compared with 13.3% when *AIM to* is used.

AIM at

"increase"		**"decrease"**		**"halting"**	
verb	**number of occurrences**	**verb**	**number of occurrences**	**verb**	**number of occurrences**
improving	29	reducing	30	ending	50
boosting	13	curbing	14	stopping	8
increasing	12	cutting	6	stabilising	3
developing	9	slashing	2	blocking	2
raising	6	showing	2	restricting	2
bolstering	2	limiting	1	halting	2
	8.4%		6.5%		7.9%

AIM to

"increase"		**"decrease"**		**"halting"**	
verb	**number of occurrences**	**verb**	**number of occurrences**	**verb**	**number of occurrences**
raise	26	cut	17	end	14
improve	18	reduce	14	stop	7
develop	15	halve	2	finish	5
promote	14	limit	2	halt	2
boost	7	minimise	2	stabilize	1
increase	6	shrink	2		
add	2	slash	1		
maximise	2				
gain	1				
	7.6%		3.3%		2.4%

Figure 5 : Verbs occurring at N+1 of *AIM at* and *AIM to* representing denotational synonyms of 'increase', 'decrease' and 'halt', no occurrences of negative polarity being found

2.2.2. Colligation

Stubbs [2001: 65] defines colligation as being

> the relation between a pair of grammatical categories or, in a slightly wider sense, a pairing of lexis and grammar.

Taking the larger sense of this definition into account and considering both grammatical words and concepts to be included, three areas of interest will be examined, namely tense frequency, relative clause expression and negation.

Firstly, one main point demonstrating a marked preference concerning colligation is the grammatical form of each lemma as presented by the returned corpus query (the number of cases for each form of *AIM at* and *AIM to* may be seen in figure 6). Of the sixteen examples of the form *aim* used with *at*, 8 are caused by a preceding modal verb including one instance of *will*, 3 denote imperative clauses, 1 is an infinitive form and only 4 are used as a declined, present tense verb form. It can thus be seen that *AIM at* attracts the form V-ed in 95% of cases in contrast to *AIM to* where only 6.9% of V-ed can be seen. The notion of V-ed forms implies that a hypothesis may be confirmed as having been successful or not due to its qualities as past tense. This may account for its high frequency of use in the media where precise information is given and confirmation of success or failure is expected. The high rate of occurrence of present tense forms in the sub-corpus **ukephem** may highlight the potential, but not yet accomplished nature of desired outcomes. The passive voice, inducing the V-ed form, may be seen 223 times with *AIM at* compared with only 7 times with *AIM to*. This may in part be due to the use of *AIM at* in a majority of sub-corpora of relatively high formality which often favour the comparatively frequent use of the passive voice.

	AIM at	
form	**number of occurrences**	**% of occurrences**
aim	16	1.9
aimed	805	95.5
aiming	8	0.9
aims	14	1.7

	AIM to	
form	**number of occurrences**	**% of occurrences**
aim	387	32.4
aimed	83	6.9
aiming	213	17.8
aims	512	42.8

Figure 6: Frequency of occurrence of individual forms of the lemmas *AIM to* and *AIM at* according to the number of occurrences and the relative frequency expressed as a percentage

Observations may also be made concerning negation as very few examples indicating negative polarity may be seen concerning *AIM at/to*. In order to compare the number of instances found, it may be useful to compare results of negation found in other studies. For

example, taking a random sample of 18 million words of the BoE corpus, Halliday and James [1993] concluded that the relation of positive to negative clauses was 9:1. In a slightly different aspect, Tottie extracted a 50.000 word sample of the Survey of English Usage Corpus and, whilst 12.8 cases of negation were detected per 1000 words in written discourse, the occurrence in spoken discourse was discovered to be more than twice as frequent, reaching 27.6 cases per 1000 words [Tottie 1991]. If Tottie's results are to be compared with those of Halliday and James, and also with those found in this study, it is first necessary to find the corresponding ratio of positive to negative clauses by estimating the number of sentences per thousand words. Using Francis and Kucera [1982]'s model of sentence length, given by Kennedy [1998: 158], based on the whole of the Brown Corpus, the average number of words in a phrase amounts to 18.4. Applied to Tottie's statistics, it may be deduced that the relation of positive to negative clauses would amount to 8:1 in written discourse, which roughly corresponds to Halliday and James' findings, and 4:1 in spoken discourse, which may however seem high. Tottie explained the differences in results as being due to the low degree of interaction characteristic of written communication compared with spoken discourse.

A study of *AIM at/to,* however, shows a highly significant difference in statistics compared with the two studies mentioned above as, when considering *AIM at/to*, the relation of positive to negative clauses can be seen to be 100:1. There appears to be no grammatical or semantic reason preventing negation with the lemma studied yet when requesting examples of *AIM* expressing negative polarity, only 19 cases are found: 10 concerning *AIM at* and 9 affecting *AIM to*, with only one appearing in the British spoken sub-corpus. This may be the result of the high proportion of cases of *AIM at/to* found in media or ephemera-based sub-corpora where in both text types goals are asserted. This surprisingly low ratio of positive to negative clauses may also be influenced by the speaker's projected involvement in the action, the semantic notion attached to *AIM at/to* expressing what is desirable from the Agent's point of view; a positive clausal polarity seems here to be adapted to this purpose. The lack of negative clauses thus appears to be influenced by other factors in the unit of meaning which may support Sinclair's notion of semantic prosody [Sinclair 1996].

2.2.3. Discourse prosody

Discourse prosody is defined by Stubbs [2001: 88] as being

> a descriptor of speaker attitudes and discourse function.

He further explains this stating that discourse prosodies are evaluative and "often express the speaker's reasons for making the utterance" [Stubbs 2001: 65], echoing Sinclair [1986]'s notion of semantic prosody where meaning is spread over a unit of meaning and not confined to any particular word.

The lack of negative clauses has already been noted in relation to *AIM at/to,* and this is further reinforced by the association of the verb at N+1 as well as collocates at N+2. The desirable or undesirable sense of the main verbal collocates of *AIM at/to* cannot be determined in many cases without reference to the surrounding units. For example, verbs such as *improve, develop* or *promote* may carry desirable semantic associations, yet others such as *end, force, cut,* or *reduce* need to be examined in their surrounding context as they may be associated with undesirable notions. Figure 7 shows twenty randomly selected Key Word in Context (KWIC) concordance lines (with added line numeration) of *AIM at/to* + *reducing/reduce*. It may be seen that the use of a verb at N+1 having potentially undesirable connotations collocates with a noun at N+2 which also possesses such associations. This collocation of two items having, individually, undesirable connotations tends, when used

together, to project a desirable goal, for example *to reducing teenage smoking* (l. 10) or *reduce emissions of carbon monoxide* (l. 16). This pattern can be seen as being repeated equally throughout the positive clauses of both *AIM at/to* and is also reflected in the negative clauses. For example, in the following line from the sub-corpus **ukbooks** the basic notion of "cultivating a sweet tooth" carries widespread undesirable connotations which can appear to be tentatively neutralized by the speaker's prior negation:

> this approach is not aimed at cultivating a sweet tooth

1	bbc/06	has donated nine-million dollars to an international programme	aimed	at	reducing	the use of chemicals known as CFCs (chlorofluorocarbons),
2	bbc/06	next month for a second round of talks	aimed	at	reducing	tension over Kashmir. <h> India has made fresh
3	bbc/06	has agreed on large cuts in public spending	aimed	at	reducing	the country's growing budget deficit, caused by unification.
4	oznews/01	2.3 million over four years to develop programmes	aimed	at	reducing	family violence including specific training for relationship counsellors.
5	oznews/01	<p> Goodman and Defiance had been in talks	aimed	at	reducing	operating costs by rationalising some of their rival
6	oznews/01	<p> The State Government had launched several initiatives	aimed	at	reducing	alcohol consumption among young people. <p> AAPINTNEWS
7	sunnow/17	plan is part of a <KPD> 22million package	aimed	at	reducing	truancy and expulsions unveiled yesterday by schools standards
8	ukbooks/08	attention can then be directed at practical policies	aimed	at	reducing	dependency. However, although it might be useful to
9	ukephem/02	part at Glasgow Dental Hospital in a project	aimed	at	reducing	the level of tartar build-up. <p> To qualify,
10	usbooks/09	an `Adult Choice." Many tobacco company programs ostensibly	aimed	at	reducing	teenage smoking tell teenagers that they are too
11	sunnow/17	Black scoring avalanche. World champions women not to	aim	to	reduce	weight drastically. Perhaps the best guideline is to
12	sunnow/17	just couldn't get it together. to save electricity	aimed	to	reduce	energy costs by a relatively-small 10 percent over
13	sunnow/17	appalling conditions. The squaddies were goods. Er they	aimed	to	reduce	the budget deficit of from twenty per cent
14	sunnow/17	when you leave Gazza at The bank is	aiming	to	reduce	its equity later this week, offering buyback on
15	sunnow/17	three-quarters full when Sam stepped SELA officials which	aims	to	reduce	foreign debt payments, mainly by encouraging countries to
16	sunnow/17	were extraordinary. We want to a plan that	aims	to	reduce	emissions of carbon monoxide by the year 2000.
17	sunnow/17	blamed for 68 deaths and by police nationally,	aims	to	reduce	the number of car accidents on roads over
18	sunnow/17	Josephs in today's DJ Laing million. <p> It	aims	to	reduce	trade barriers in goods and services, promote investment
19	sunnow/17	club. They have already drafted J Group	aims	to	reduce	sports injuries By JAMES SWANWICK AN Australian sports
20	sunnow/17	up to 500 jobs. The PBF, a charity,	aims	to	reduce	the financial impact of paralysis at the time

Figure 7: KWIC concordance lines for *AIM at/to+reduce/reducing*

2.2.4. Distribution in text types

Before looking at statistical data on the distribution of *AIM at* compared with *AIM to* in various text types, it is first necessary to define the field concerned. For Biber, text types have a strictly linguistic basis being "sets or groupings of texts such that the texts within each set are linguistically similar while sets are linguistically distinct" [Biber and Finegan 1991: 213]. Types are determined by their predominant linguistic features and given functional labels. Readily distinguished categories such as newspapers, novels or broadcasts are considered to be genres. Biber's definition of genres conforms to the notion of text types as defined by Stubbs [2001: 20] who states that

> different text types have different patterns of expectation.

It is this description of text types which shall be considered in this study.

It has been recognised that grammatical categories, vocabulary, and semantic patterns vary according to text type. The distribution of the lexical units *AIM to* and *AIM at* may be seen to follow these predictions as can be noted in the representation of average occurrence in sub-corpora shown in figure 8. In the CobuildBoE corpus, *AIM at* is used with the greatest frequency in the sub-corpus **bbc** which accounts for 36.1% of its total occurrence in all sub-corpora. This strongly contrasts with the use of *AIM to* in the same sub-corpus, accounting for only 7.5% of its use throughout all the sub-corpora. This pattern is reflected less markedly in the highest use of *AIM to* seen in the sub-corpus **ukephem** where the lemma's occurrence accounts for 17% of its total corpus use. Parallel to this, *AIM at* in this sub-corpus amounts to a mere 4.8% of its use throughout the CobuildBoE, thus emphasising preference for one specific form for a given text type.

	AIM at	***AIM to***
sub-corpus	**% average/million words**	**% average/million words**
bbc/06	36.1	7.5
npr/07	16.7	13.6
oznews/01	14.3	0.8
sunnow/17	3.7	10.8
times/10	7.9	14.2
today/11	5.1	12.2
ukbooks/08	3.3	5.2
ukephem/02	4.8	17.0
ukmags/03	3.2	13.6
ukspok/04	0.4	1.5
usbooks/09	2.2	1.9
usephem/05	2.2	1.8

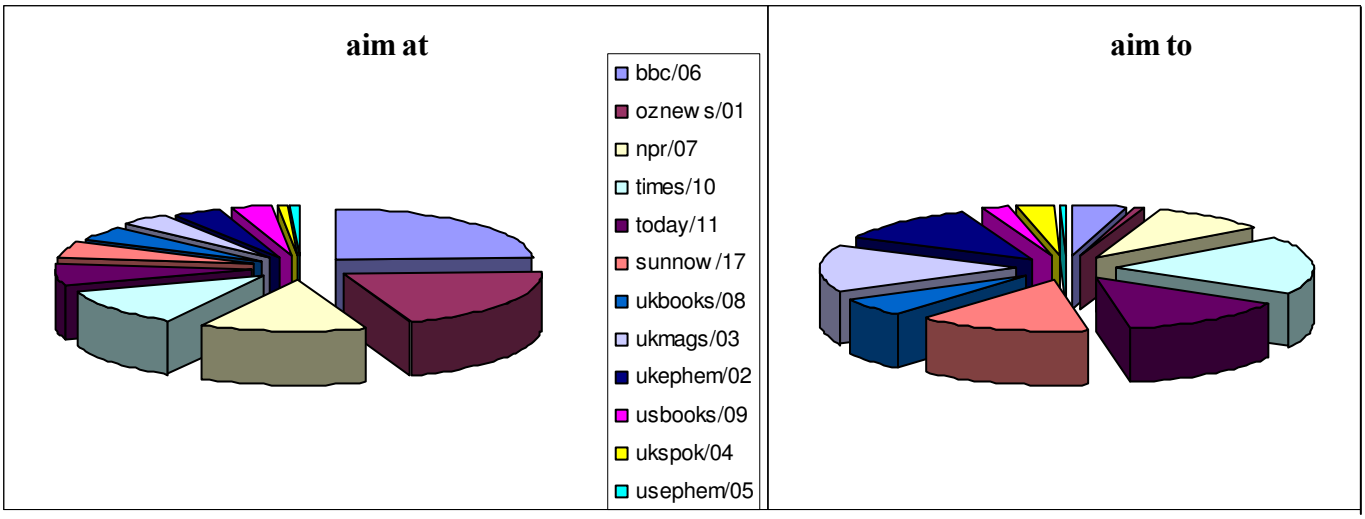

Figure 8: The distribution of *AIM at* and *AIM to* in each sub-corpus expressing the average number of occurrences per million words as a percentage

It may also be noticed that the lemma *AIM at* appears with significantly relative frequency in **oznews** and **npr,** showing the high representation of *AIM at* in radio and newspaper journalism. The sub-corpora **sunnow** and **ukmags** are high in the frequency order of *AIM to* in contrast to their number of occurrences of *AIM at*, possibly due to aspects of N+1 collocates, as shall be seen below. One sub-corpus where all forms of the lemmas appear rarely is **usephem** which only accounts for roughly 2% of occurrences in both cases; this may seem surprising if compared with **ukephem** especially in the case of *AIM to*.

Collocations may also appear significant if examined according to text type. This is, not surprisingly, the case with sub-corpora with marked tendencies to favour one form in particular. If collocates appearing 10 times or more in any one sub-corpus are selected, certain lexical preferences may be seen. Searching all sub-corpora, 60% of the use of the verb *ending* as immediate N+1 collocate may be found in the sub-corpus **bbc,** 70% of the use of *resolving* and 30% of *reducing*. With *AIM to*, the sub-corpus **ukephem,** for example, appears to favour the use of *provide* (50% of use as a collocate) and *help* (28% of use). The sub-corpus **sunnow** appears to favour more semantically general collocates such as *make*, accounting for 40% of its use in N+1 collocation. This is coherent with evidence found of the high occurrence of *AIM to* in this sub-corpus as well as the marked preference shown by the lemma for more semantically general verbs at N+1.

2.3. General discussion of results

It may be seen from corpus evidence that there are no notable semantic differences between *AIM at* and *AIM to* and thus other factors may lead to the preference of one form rather than the other. Drawing on all aspects treated, several hypotheses may be proposed.

The relatively high frequency of *AIM at* has been noted in radio corpora, especially **bbc**, thus leading to possible conclusions as to the predictability of occurrence of *at* rather than *to*. The relatively high level of perceived formality of corpora such as **bbc** or **npr** may partly be responsible for this choice due to aforementioned collocational considerations, whether to the left or right of the node. Indeed, precise and formal N-1 collocates appearing simultaneously with verbs possessing precise semantic meaning at N+1 seem to indicate two linked tendencies: firstly, the high probability of such a combination appearing in sub-corpora such

as **bbc, npr,** or **oznews** and secondly, the high number of occurrences of *AIM at*. This is a stark contrast to observations made when considering ephemera, magazine or tabloid corpora (**ukephem, ukmags, sunnow**) where the occurrence of *AIM to* largely predominates. In such corpora, less semantic precision may be noted whether at N-1 or N+1 and the level of perceived formality may be considered as being low.

The tense of each form of the lemma may also be considered of importance. The high majority of the form V-ed used with *AIM at* has already been discussed yet reasons for this may be proposed by considering the general pattern of V-ing clauses as complements which express actualisation contrasting with the notion of to-infinitive clauses which indicate potentiality. This latter may result in the majority of present tense use seen with *AIM to*, as the Agent expresses present desires which may or may not be actualised, the unsure outcome not justifying the choice of a past tense; this may especially justify the frequent use of AIM to in **ukephem** where the potentiality of actions is emphasised.

The absence of the lemma in all its forms in spoken discourse may also be noted, leading to conclude on the possibility of preference for more semantically general verbs or units such as *want* or *would like to*.

The results given have been found using only the BoE. However, further scope may be given to this issue by examining in greater detail the structure "it is aimed at/ it is aimed to" by looking it up through a readily available web-based search engine. Using Yahoo, for example, one notices that the structure 'AIM to' is much more common than 'AIM at' with a ratio of over 90% to 10%, mainly because "it is aimed at" accepts both a noun (people, children…) or a "V-ing" structure while "it is aimed to" is almost always followed by a verb and hardly ever by a noun.

From a stylistician's point of view, the patterns revealed by corpus analysis may be revealing, as they often determine the impact of a phrase. The seemingly natural tendency, for example, of using more semantically precise lexical items as collocates of *AIM at*, is an indication of the register of language desired by an author. Also, the observation that both lemmas tend not to collocate with occurrences of negative polarity may draw attention to the exceptions.

The main area of overlap between a corpus-driven study and stylistics may be seen in applications to discourse prosody. For example, the stylistician may note in an individual text that collocates at both N+1 and N+2 positions may individually have undesirable connotations, but when used together become a rather positive goal (see examples 2.2.3). What may be of generally more use to the stylistician is the possibility given by corpus tools to indicate that this is a norm, thus allowing deviations from this norm to be highlighted.

Conclusion

Despite few distinctions being made in dictionaries, *AIM at* seems to follow different patterns of use compared with *AIM to*. The corpus evidence may be seen to support Stubbs' idea that data confirms dictionary examples without giving underlying principles. This may be applied to the frequent use of the V-ed form with *AIM at*; thus, in a dictionary such as CCELD, it may be useful to mention that this is the most usual form met. Two equivalent examples, namely *AIM at/to* + VERB, should also be given to reduce confusion.

Certain points or typical patterns which may not be obvious when comparing relatively few examples of such a word form may be observed more efficiently using corpus techniques, confirming the idea formulated by Stubbs [2000] that "the crucial shift is from studying what is possible to what is probable". From a learner's point of view, it may also seem important for words to be studied according to surrounding context and not as isolated units.

The description of patterns and highlighting of repeated occurrences may therefore be of use not only to the lexicographer, but also to the stylistician. By determining typical usage with large amounts of data, deviation from this norm may be easier to spot. In a broader scope, corpus tools compiled by and for the lexicologist may prove to be a useful resource for the stylistician in all areas of stylistics, and not only where lexical matters are concerned. Continuing technical advances and the growth in available resources may hasten further the sharing of the lexicographer's toolkit with mainstream stylistics.

Bibliography

AARTS J., "Corpus Linguistics: an appraisal", Paper read at the fifteenth International Conference on Literary and Linguistic Computing, Jerusalem, June 1988, in AJIMER K. & ALTENBERG B., *English Corpus Linguistics*, New York, Longman, 1991: 45-46.

AARTS J., "Intuition-based and observation-based grammars", in AJIMER K. & ALTENBERG B., *English Corpus Linguistics*, New York, Longman, 1991: 46.

BARNBROOK G., *Language and Computers*, Edinburgh, Edinburgh University Press, 1996.

BIBER D. & **FINNEGAN** E., "On the exploitation of computerized corpora in variation studies", in AJIMER K. & ALTENBERG B., *English Corpus Linguistics*, New York, Longman, 1991: 213.

DE ROSE S., "An analysis of probabilistic grammatical tagging methods", in JOHANSSON S. & STENSTRÔM A-B., *English Computer Corpora*, Berlin, Mouton de Gruyter, 1991: 9-14.

FRANCIS W. N. & **KUCERA** H., *Frequency Analysis of English Usage: Lexicon and Grammar*, Boston, Houghton Mifflin, 1982.

HALLIDAY M. A. K. & **JAMES** A., "A quantitative study of polarity and primary tense in the English finite clause", in SINCLAIR J., HOEY M. & FOX G. (eds), *Techniques of Description: Spoken and Written Discourse*, London, Routledge, 1993: 60.

ITKONEN E., "Qualitative vs quantitative analysis in linguistics", in PERRY T., *Evidence and Argumentation in Linguistics*, Berlin, de Gruyter, 1980: 344.

KENNEDY G., *An Introduction to Corpus Linguistics*, Longman, 1998.

MAHLBERG M., "Corpus stylistics: methodology, theory, and patterns in literary texts", 2006: http://ahds.ac.uk/litlangling/events/approaches/mahlberg.htm, accessed 10/02/10.

POTTER L., *A Guide to Using the CobuildDirect Corpus*, Birmingham, University of Birmingham, 1999.

SINCLAIR J., "Lexicographic evidence", in ILSON R. (ed.), *Dictionaries, Lexicography and Language Learning*, Pergamon and the British Council, 1985: 81-94.

SINCLAIR J., "The search for units of meaning", *Textus*, 9, 1996: 79-106.

SINCLAIR J., "The lexical item", in WEIGLAND E. (ed.), *Contrastive Lexical Semantics*, Amsterdam, Benjamins, 1998: 1-24.

SINCLAIR J., *Corpus, Concordance, Collocation*, Oxford, Oxford University Press, 1991.

STUBBS M., "Society, education and language: the last 2,000 (and the next 20?) years of language teaching", in TRAPPES-LOMAX H. (ed.), *Change and Continuity in Applied Linguistics*, Clevedon, BAAL and Multilingual Matters, 2000: 15-34.

STUBBS M., *Words and Phrases: Corpus Studies of Lexical Semantics*, Oxford, Blackwell Publishers, 2001.

STUBBS M., "Conrad in the computer: examples of quantitative stylistic method", *Language and Literature*, 14: 1, 2005: 5-24.

TOTTIE G., *Negation in English Speech and Writing*, London, Academic Press, 1991.